DOCUMENT RESUME

ED 370 967 TM 021 516

AUTHOR Resnick, Lauren B.; Resnick, Daniel P.
TITLE Issues in Designing and Validating Portfolio Assessments. Project 2.3: Complex Performance Assessments: Expanding the Scope and Approaches to Assessment.
INSTITUTION Center for Research on Evaluation, Standards, and Student Testing, Los Angeles, CA.; Pittsburgh Univ., Pa. Learning Research and Development Center.
SPONS AGENCY Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE Nov 93
CONTRACT R117G10027
NOTE 39p.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Educational Assessment; Elementary Secondary Education; English; Evaluation Methods; Mathematics; Measurement Techniques; *Portfolios (Background Materials); *Scoring; *Student Evaluation; Teaching Methods; *Test Construction
IDENTIFIERS *New Standards Project (LRDC); *Performance Based Evaluation

ABSTRACT

Portfolios, with their associated exhibitions, will be the heart of the New Standards assessment system that is being developed. It will be necessary to combine the functions of portfolios as measurement tools and as tools for instruction and learning. As a beginning, the New Standards project has examined how teachers are already using portfolios through a task development meeting and has considered implementation issues and approaches to scoring at a 1993 meeting for teachers and curriculum supervisors in mathematics; a similar meeting was held for teachers of English. In September 1993 New Standards began a design process that will include a meeting of 42 teams of teachers from around the country who will be developing portfolios in their classrooms in 1993-94. Attachments to this document include the following: (1) "Environmental Scan, New Standards Project, Portfolio Study" (New Standards Project); (2) "Mathematics Portfolios, March 1993" (New Standards Project); (3) "Issues in Scoring Cumulative Accomplishments: Implications for Portfolio Design. A Paper Prepared for the New Standards Project" (Philip Daro); and (4) "New Standards Takes a Closer Look at Portfolios" (newspaper reprint). Two figures illustrate the discussion, and a listing of the New Standards Portfolio Development Teams is included. (Contains 1 reference.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 370 967

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

National Center for Research on
Evaluation, Standards, and Student Testing

Final Deliverable – November 1993

Project 2.3: Complex Performance Assessments:
Expanding the Scope and Approaches to Assessment

Issues in Designing and Validating Portfolio Assessments

Lauren B. Resnick and Daniel P. Resnick, Project Directors
CRESST/LRDC, University of Pittsburgh

U.S. Department of Education
Office of Educational Research and Improvement
Grant No. R117G10027 CFDA Catalog No. 84.117G

Center for the Study of Evaluation
Graduate School of Education
University of California, Los Angeles
Los Angeles, CA 90024-1522
(310) 206-1532

BEST COPY AVAILABLE

TM021516

The work reported herein was supported in part under the Educational Research and Development Center Program cooperative agreement R117G10027 and CFDA catalog number 84.117G as administered by the Office of Educational Research and Improvement, U.S. Department of Education.

The findings and opinions expressed in this report do not reflect the position or policies of the Office of Educational Research and Improvement or the U.S. Department of Education.

ISSUES IN DESIGNING AND VALIDATING PORTFOLIO ASSESSMENTS

Lauren B. Resnick and Daniel P. Resnick

**CRESST/Learning Research and Development Center,
University of Pittsburgh**

Portfolios, with associated exhibitions, will be the heart of the New Standards assessment system. In developing the portfolio program, it will be necessary to combine two functions of portfolios: portfolios as measurement tools and portfolios as tools for instruction and learning. When portfolios are developed primarily as learning tools, educators focus on how to establish portfolio "cultures," in which students develop important projects and learn to judge and critique their own work (Wolf, 1989). Portfolio cultures call for much student responsibility for what to work on and what to put forward as "best work." Those who have invested effort in developing portfolios as instructional vehicles are concerned that outside scrutiny could interfere with the emergent and delicate process of self-analysis and local judgment. On the other hand, use of portfolios for measurement purposes calls for some degree of standardization. Common frameworks, common scoring criteria, and reliable ways of applying them will be required. It may also be necessary to include some common tasks or task components in students' portfolios.

The measurement and learning functions of portfolios can interfere with one another if we are not thoughtful and sensitive in designing our system. An important first step for New Standards was to find out how teachers in our partner states and districts were already using portfolios. At a task development meeting held in Dallas, Texas, on November 12-15, 1992, Lizanne DeStefano conducted individual and table interviews with 50-60 lead teachers. Interviews focused on current instructional practices and factors that are likely to facilitate or interfere with portfolio implementation. The following month, site coordinators from all of the partners were surveyed by mail and telephone. The results of these interviews and surveys are reported in the attached document "Environmental Scan: New Standards Project Portfolio Study."

Our next step was to consider implementation issues and pose possible approaches to portfolio scoring based on common standards. In March 1993, we convened a two-day portfolio meeting in Charleston, South Carolina. Approximately 40 mathematics teachers and curriculum supervisors from partner states and districts with existing portfolios programs attended the meeting. Much of the first day was devoted to learning about the portfolio programs represented. Reports were given by representatives from Vermont, Pittsburgh Public Schools, Kentucky, California, and Washington (see attached "Mathematics Portfolios"). On the second day of the meeting, participants formed work groups and discussed issues surrounding portfolio implementation. The issues raised at the Charleston meeting were carried back to the New Standards management team for future consideration. The paper "Issues in Scoring Cumulative Accomplishments: Implications for Portfolio Design" emerged from several discussions of New Standards management team and advisory committees (see attached).

A second portfolio meeting was held August 11-13 in Minneapolis, Minnesota. Approximately 30 people attended, including subject area researchers, curriculum supervisors, and teachers of English language arts. All participants were connected to a portfolio network of some kind. Much of the agenda was devoted to presentations on the various portfolio networks. The latter part of the meeting focused on similarities and differences across programs and recommendations for New Standards implementation. The attached newsletter article provides additional details on the meeting (see "New Standards Takes Close Look at Portfolios").

In September 1993, New Standards began a design process for portfolios that blends the expertise of researchers and teachers, whose orientation is primarily toward the instructional function of portfolios, with the technical expertise of psychometricians and other measurement specialists. Working with leaders from education reform projects (e.g., Education Development Center, Balanced Assessment, Harvard Project Zero, Performance Assessment Collaborative for Education) and states and districts with experience using portfolios (e.g., Vermont, Kentucky, Fort Worth), New Standards has begun co-developing a portfolio system that will incorporate the best current thinking about portfolios as instruction and learning vehicles and

will produce a scoring system capable of judging various portfolio configurations against a shared set of explicit standards.

Forty-two teams of teachers from throughout the U.S. have been invited to participate in a process aimed at producing clear statements of what elements must be included in a portfolio, a library of exemplars of portfolio entries, and rubrics for scoring individual elements and full portfolios (see attached listing of New Standards Portfolio Development Teams). These will be organized into *Portfolio Handbooks* written for students, with accompanying versions for teachers and parents. The handbooks will explain the criteria and processes for assessment, give suggestions for selection and preparation of portfolio submissions, provide self-assessment tools, and offer a library of annotated exemplar portfolios.

The mathematics portfolios will include selections of problems, investigations, "gap fillers," and reflective statements by students and teachers. These are illustrated and described in Figure 1. The scoring criteria and process will be based on the NCTM Standards, interpreted by the *Mathematics Framework for Balance*. An example of a score profile that might be used is given in Figure 2. The English language arts portfolios will be similar in nature. The 42 portfolio teams, which are convened in working meetings three times per year, are developing portfolios in their classrooms in the 1993-94 school year. These portfolios-in-progress are the basis for empirical development of the criteria and scoring procedures and of the handbooks that communicate these to students, teachers, and parents.

Reference

Wolf, D. P. (1989). Portfolio assessment: Sampling student work. *Educational Leadership, 46*, 4-10.

Environmental Scan
New Standards Project
Portfolio Study
Dallas, Texas
November 12-15, 1992

The purpose of this inquiry is to increase our understanding of the process of constructing, scoring, and using collections of student work to evaluate individual students and entire schools. It is hoped that what is learned will inform the design of the portfolio component of the New Standard assessment system to be piloted in the 1993-1994 school year.

Because we hope to implement a portfolio system among New Standards Schools beginning in 1993, it is important to understand the context in which this system will operate, to assess the capacity of teachers and others who will be responsible for implementing the system, and to identify factors within schools that are likely to facilitate or interfere with implementation. For that reason, we are talking with teachers, administrators and site coordinators to get their sense of what constitutes a viable portfolio system and implementation strategy. For this report, we completed individual and table interviews with 50-60 teachers at the task development meeting in Dallas in November. We also sent surveys to site coordinators in all 21 partners. Some of these were completed by phone, some were filled out by the site coordinators and mailed or faxed back to us. In all, we received replies from 16 partners. Responses of both teachers and partners are summarized below.

Capacity of teachers and others who would be implementing the system

How would you describe a portfolio assessment system?

The majority of teachers saw portfolios as a collection of student work that might be used to document progress and to reflect on work with students and parents. Portfolios were seen as particularly helpful for accommodating complex and lengthy projects, for allowing student's to represent learning in alternate ways, for documenting change over time and for explaining to others outside the classroom the types of activity a student engaged in and how he/she performed. Teachers regarded the collection of student work over an extended time period as an important dimension of portfolios. The length of time varied for a unit to a grading period, to a semester. Only a few teachers had experience with a year long portfolio.

The primary purpose for portfolios cited by teacher was individual reflection and self assessment. Only five of almost 50 teachers interviewed viewed portfolios as a means of representing opportunity to learn or in any high stakes way such as judging students or classrooms. The was serious concern about high stakes uses supplanting the more individual uses. Teachers felt that in order to develop the link between assessment, good instruction and high standards, students and their

teachers had to be actively involved in assembling and evaluating portfolios. Teachers valued local autonomy with regard to selecting specific items to be included in the portfolio. They desired guidelines and exemplars for the process of establishing a portfolio system in their classrooms.

Few teachers had experience with scoring or formally evaluating portfolios. The ones that did cited both theoretical and practical challenges associated with the processes of judging the quality of a portfolio. A few who had been involved in large scale scoring talked about the reality of shipping portfolios from one part of the state to another, how things got lost or smashed, the cost of efficient mailing, and problems with confidentiality. One person raised issues of verifying how much feedback or assistance was given to students when preparing the portfolio when comparing across classes.

Site coordinators views of portfolios tended to focus on logistic and technical aspects. About half saw the greatest impact of portfolios at the classroom level as a tool to drive instruction and to directly involve students, teachers, and parents in the educational process. In fact, four out of 16 site coordinators strongly resisted the idea of high stakes portfolios for students, preferring that portfolios be used for individual reflection and local improvement alone. The remainder of the respondents valued the use of portfolio information that would be aggregated in some way, evaluated in terms of agreed upon standards and used to represent the quality of instruction as well as student performance. In many cases, they viewed the portfolio component as a necessary complement to other state or district assessment initiatives.

The overwhelming opinion of both teachers and site coordinators (but more strongly for teachers) was that while some elements of portfolios might be standardized for a district or state or across partners, what goes into a portfolio should be largely a local decision (otherwise it would not truly reflect local curricula). Composition of portfolios should be guided by a set of general rules and timelines, with lots of autonomy residing at the local level. Some sample guidelines that were suggested:

> Students should evaluate and annotate their own portfolios at regular intervals and at the end of a unit or school year. Student evaluations should become a part of the portfolio.
>
> Standards for portfolios should be explicit and available to parents, students, and teachers from the beginning. In fact, they should be involved in setting the standards.
>
> Parents should be asked to review and evaluate their child's portfolio.

A panel made up of parents, teachers, administrators and community members should design or select the number and types of activities to be included in portfolios for particular grade levels.

The NSP should set the standards and let states and districts worry about what to put into the portfolio to meet them. You would be surprised how creative teachers can be when challenged in that way.

Most teachers and site coordinators frequently saw portfolios as containing organic components such as selected classroom work, group projects, and individual projects and extended work. Often, but less frequently, mentioned were the inclusion of curriculum imbedded tasks. On demand tasks were not often cited as part of a portfolio.

What types of experiences are teachers in your school/district/state likely to have had with portfolios or other cumulative assessment systems?

Experiences with portfolios among teachers as described by site coordinators and teachers themselves varied widely. It is safe to say that teachers have had more experience with portfolios in reading than in math. Secondary teachers had more experience that elementary teachers. Virtually all portfolios that were mentioned were unidisciplinary. Independent of their level of experience, teachers were generally enthusiastic about portfolios and their use in NSP.

Every site coordinator reported that something was going on in their state with regard to portfolios. Vermont and Kentucky have formal and articulated mandated statewide systems. California, Maine, and Oregon have had extensive implementation in demonstration sites, but both Oregon and Maine emphasize local rather than other types of use. Other state have smaller demonstrations or individual districts/schools/teachers opting to use portfolios. It should be noted, that even in states with mandated systems, teachers and site coordinators were quick to say that a great deal of variance existed. I think it is fair to say that except in the case of the states with mandates, most teacher participation has been voluntary, of short duration, and classroom based. Other than the statewide systems, few partners reported experiences with portfolios that included standards for assembly or scoring or that carried with them high stakes consequences.

--How much variance exists? Describe the extremes.

The complete range, from no knowledge to full scale implementation. The mode is probably first hand, time limited experience collecting student work for instructional purposes within a single content domain.

What are reasonable expectations in terms of time demands of a portfolio system?

Teachers were concerned about both time and the physical management of portfolio contents. Site coordinators were concerned about scoring time and costs. Both were very interested in exploring high tech options for management and scoring such as computer scanning and videos. All groups were reluctant to state a specific amount of time that teachers might devote to the assembly and scoring of portfolios. They seemed to feel that if the system were of high quality, closely aligned with state and NSP goals, and valued–time would not be an issue. It was clearly expressed, however, that if portfolios were added, something had to be taken away. This might mean reduction in standardized test administration, documentation of progress in basal tests, or individualized progress reports. It seems to suggest that when a local site decides to implement portfolios, a conscious and public decision should be made that some parallel procedure will be eliminated. Teachers at model sites said that this was a motivator to get teachers to begin to use portfolios. In one case, teachers were given the option of using a cumbersome system of unit tests in a basal reading series or a portfolio system. The portfolio system appeared easier to implement, more instructionally relevant and interesting. Teachers quickly chose to participate.

What types of professional development or information would be necessary to get a portfolio system up and running in your locale?

Not one person interviewed felt that professional development could be handled by time limited inservice alone. Most interviewees voiced the need for intensive, long term training such as 8 week telecourses, four week summer workshops, and 50 hour certification programs for at least a few local staff who could then be relieved of teaching or supervision responsibilities to train others. The importance of including principal and supervisors in the training was stressed. Respondents felt the need for onsite support as well. They suggested that demonstration sites might be clustered near a university or some other source of technical assistance or that NSP personnel should be hired to provide regional TA. Public engagement and education for parents and community members was a need cited by over half of the site coordinators.

--What resources are available to provide this training?

Site coordinators and teachers identified limited resources from federal and state departments of education, universities, and consultants. Most teachers who had participated in training described half or full day sessions during teacher inservice days. All agreed that resources would have to be greatly increased to respond to these initiatives. There was no consensus as to where these additional resources would come from. Most felt that their districts or states did not have the magnitude of resources that the effort demands

Barriers and enhancers to a portfolio system

What are some factors within your state/district/school that might facilitate the implementation of a portfolio system?

Respondents cited the generally positive attitude that teachers have toward portfolios as the main enhancer to implementation. It was expressed that portfolios of some sort have been used by teachers for years, that the idea makes sense to them, and that the information to be gained from portfolio review is highly useful. The wave of interest in portfolio assessment and dissatisfaction with standardized testing were also cited as reasons why some schools may choose to implement a portfolio system.

--Inhibit implementation?

Comments indicated a strong need for public and professional engagement. It was commonly thought that standardized test scores were the only information that school boards and legislators found credible. Respondents felt that if this opinion could be changed, resources and commitment to portfolios would be forthcoming.

Logistic issues were also cited as barriers to participation in a portfolio system. Some teachers said that they had difficulty getting file cabinets and other material needed to start a portfolio system. Others felt that the organization of secondary schools, when one teacher might see 170 students per day made any widespread system physically impossible. Administrators cited the difficulty in transferring school records from one school to another when a student moves. They felt that the use of portfolios exacerbated the problem.

Technical issues were also raised. Respondents were worried about equity of a portfolio system, given the distribution of high quality instruction and resources in our schools. Respondents questioned the rewards and consequences of reporting and using portfolio information for students, for teachers, and for schools. They felt that the motivational aspects of the portfolio system and its consequences should be apparent and strong and that scoring should be sensitive to both individual and institutional growth.

Other technical concerns were related to issues of validity, generalizability and moderation. In the opinions of the respondents, technical issues associated with making and comparing judgments of different types of work by different students from different classrooms have not begun to be addressed. It was generally felt that a key part of this complex situation would be the development of a set of widely agreed upon and broadly applicable set of performance standards for judging quality work.

Interaction with other assessment initiatives within the state

Do current assessment initiatives in your school/district/state include a portfolio component? What is it like?

See first question.

--Could it serve as an example of best practice?

Vermont, Maine, and Oregon nominated sites that they feel exemplify best practice.

MATHEMATICS PORTFOLIOS

New Standards Project March, 1993

On March 18 and 19, the New Standards Project held a meeting to begin discussing important issues around the topic of portfolio assessment. Phil Daro led the group of representatives from the different partner states who convened at Charleston, South Carolina.

As a way of introduction Phil emphasized the understanding that Portfolios are not a "quick fix." He also made clear that this meeting was not to make decisions, but to help NSP focus on issues and to gather information about what is known, being tried, and believed about portfolios. As he put it: "This is how we begin charting a portfolio course for the New Standards Project."

During the meeting partner states already using or developing a portfolio assessment system had the opportunity to share their findings and thoughts. Following are summaries of those presentations.

Vermont's Portfolios

Sue Rigney
Marge Petit
Beth Hulbert

Overview
A portfolio–is a collection of purposeful work; records of progress and achievement collected over time. It includes formative and summative information, communicates to students, teacher, and parents that assessment is for everyone, everyday. Represents self-evaluation or reflection that helps us set goals and paths to learning.

Portfolios are characterized by the fact that students, teachers, parents work collaboratively and that students see that they get better over time (they foster a positive disposition towards math).

History:
Portfolios in Vermont derived from the NCTM Curriculum Standards.
Vermont developed its general rubric around assessing that which has been underassessed (NCTM Standards 1–4: processes):

- problem solving
- communications
- reasoning
- connections

Philosophically, Vermont made a choice to develop and use a generalized rubric. The use of a general rubric is the strongest part of the system because it gets at those hard-to-assess, higher level process components.

There are efforts still being made searching for the reliability factor to satisfy all the interested parties.
The most positive aspect is that the multi-dimensional rubric clarifies and gives ownership of the process to the student —what raises the level of performance is students' knowing what is mathematically valued.

How it Works
The emerging result:
The individual student is assessed (and is used to being assessed) on the processes.

The program is assessed on content area, empowerment, and instructional opportunities.

The actual portfolio contains these things:

- a letter to evaluator
- 5-7 student selected pieces (individual, process-focused)
- 10-20 teacher selected pieces (program, content-focused)

There are seven criteria for student self-assessment:

PROBLEM SOLVING
- understanding the task
- how you solved the problem
- decisions along the way
- "so what"--outcomes of activities

COMMUNICATION
- mathematical language
- mathematical representation
- presentation

A score is given for each of the seven problem solving and communication criteria (if seven pieces, 7 scores per criterion = 49). An average within each criterion (not across criteria) yields 7 scores.

Scoring Issues
Vermont model was that teachers score their own students' work, with selected pieces audited by larger separate entity. The Rand Study resulted in redesigned training, and changed reporting The goal is a clear, consistent, and public strategy.

It has been found that the left to right "thermometer" fill in of each criteria is useful for public reporting.

The experience of 5 people scoring innumerable portfolios the first year showed that teachers often had little or no idea what was meant by "problem solving."

Several efforts are on the way in Vermont to improve the process:

1) Summer institutes:
Model has been redesigned; scoring now goes in a different pocket . Rather than focusing on scoring, the Institute's primary purpose is supporting teachers in learning new mathematics and learning to teach it.

2) Identified teachers "on-call" for portfolio ideas and professional development

3) Vermont EdNet system has helped in developing a network among teachers.

4) Getting Started: The Vermont Mathematics Portfolio was pioneered by Marge Petit/Beth Hulbert/Bill Thompson.

Vermont's big ideas

- **The generalized, multi-dimensional rubric is a strength**
- **Students and teachers learn the dimensions for self assessment**
- **Portfolios generate professional community but need professional development**
- **Teacher support is essential**

Kentucky's Portfolios

Cheryl Tibbals:

History
There was a court case mandating the re-creation of educational system in Kentucky which meant starting at "ground zero."
The main goals for 1996 are:
- restructuring
- school site councils
- ungraded primary

The components of the transitional assessment are:

1) multiple choice standardized testing for chapter One, Seven open-ended items (15 minutes items) and 1 hour writing.

2) performance events (on demand) where students perform on different "stations" in the following configuration:
at the school
a sample of students
two parts: group activity,
individual written response

3) portfolios (a collection of "best" works).

The purposes for using portfolio assessment are to:
- drive instructional changes
- help students to become self-sufficient learners
- provide two-way feedback system between student and teacher
- provide assessment which matches good instructional practices
- be used for accountability

Kentucky's view of change is generational: 20 years. This presumes incremental units of expected change. Meeting a two-year goal means a pat on the back. Exceeding it means a cash award. Not meeting a two-year goal means increasingly serious sanctions according to how far below goal (or the baseline) teachers slip.

How It Works

Decision-making has moved to the school level. Assessment, however, is required and acts very much as a lever.

The students must select 5-7 tasks (on their own, not with the teacher)

The Portfolio reflects Kentucky's view of learning and the discipline:

- Categories for entries reflect the breadth inherent in the NCTM Standards
- Students become active mathematicians
- Students evaluate their own work
- Students go beyond the right answer and have a vision of mathematics

Portfolios provide evidence of

- breadth of conceptual coverage (not textbook coverage)
- level of student "performance", level of student effort expended
- extent to which student is achieving self-sufficiency
- teaching strategies and modalities used in the instructional program
- extent to which technology and tools are used to accomplish tasks

The "Kentucky Mathematics Portfolio Teachers' Guide" was created by a task force of Kentucky math teachers. Student tasks are provided by classroom teachers within a framework of entries contained in that guide

Scoring is accomplished by:

- The Department of Education and testing contractor train regional coordinators/cluster leaders.
- Cluster leaders train teachers within their clusters.
- Classroom teachers score their own portfolios.
- An auditing system is used to insure that scoring is consistent across the state.
- Moderation sessions bring "discrepant" teachers into alignment.

Professional development issues and questions

The implementation of this portfolio assessment system raises very important questions:

What are the desired outcomes?
Why use portfolios?
How to implement as part of instructional program?
What are appropriate tasks?
What are the roles of student, teacher, parent in this process?
How much/what kind of help can teachers and parents provide?
How to get kids engaged?
How to know that scoring is aligned?

Other Comments

So far, the stakes for students are a grades. Assessing at "high school", rather than "12th grade" helps alleviate reluctance on the part of seniors. One proposal is that a certain number of "proficient" scores be required for graduation. Principals and school staffs are using portfolios at all grade levels because they see the advantage of doing so in this high-stakes setting.

Portfolio scores are available to students and teachers in October of the following school year. Working folders can have many pieces of work (even from previous years), but the scored portfolio only has the 5-7 pieces at any given time.

Right now, there's a varied range of portfolio tasks within a school or a classroom.

California's Portfolios

Part One

Nanette Seago

Setting The Stage

Even though there isn't a state-wide portfolio system in place yet, many individual teachers and/or projects have started using portfolios for assessment in classrooms and schools across the state.

These documents have provided the basis:

NCTM Curriculum Standards
California Mathematics Framework
NCTM Teaching Standards

But teachers are looking beyond this, asking the question: What are the cross-curricular, bigger ideas? Some of these include:

Growth over time
Biographies of work over time

Evidence of revision/revisiting

Evidence that this is an interactive process (finding a multiplicity of ways for students to show what they know)

Students' understanding of the purpose of compiling a portfolio of work

The participants looked at a number of portfolios which represented a wide variety of collections of work, students, and teachers. These portfolios are very much "works in progress" and provide a window onto students' understanding. Therefore they contained a wide range of work, some graded some ungraded (according to individual teachers' needs and understanding of the process). The portfolios included a range of grade levels from kindergarten to high school.

After the participants had a chance to look at the portfolios the following comments were made:

Question: ***What was revealing about the student portfolios?***

—the way the teacher perceived the portfolio task was varied.

—sometimes there was just a number (score), with mathematically incorrect work.

—there seemed to be two definitions of math (beginning and end).

—the portfolio included a variety of tasks.

—the tasks were not always physically present (this was tormenting).

—a table of contents or similar structure was appreciated (it helped the reader).

—I didn't understand the part that looked like worksheets.

—it was interesting to look at teacher's process, as well as students'.

—the work in the portfolios gave a sense of students' attitude, aside from mathematics

—it was interesting to see reflections about processes, best work.

—only a few had student reflections on every piece (these were useful)

—teacher comments back to students ranged from posing further questions, to number scores, to grades.

Question: ***What were the commonalities?***

—students communicating with diagrams, pictures

—evidence that teachers expected students' written communication

Part Two

Barbara Storms (ETS)
Karen Sheingold (ETS)

The Educational Testing Service (ETS) has been given the task of developing curriculum embedded and portfolio assessment systems for the state of California. Curriculum embedded and portfolio groups currently at work consist of teachers creating assessment tasks.
The state-mandated assessment in California—long known as CAP (California Assessment Program), has recently been "re-acronymed" to CLAS: California Learning Assessment System.

CLAS's goal is not only to provide individual results, but to build self-esteem and create lifelong learners.

Some part of the assessment may still be on demand. "On-demand" means shrink, wrapped in a testing window, but on-demand will have multi-day and may have group work components.

Curriculum embedded groups—
Four subject area curriculum embedded groups are working within the format "curriculum embedded" to answer two basic questions:

What does it look like?

What are the sources, methods, tools, circumstances of assessment appropriate to this format?

Portfolio group—
Their intent is

- to build on what has been done and is known about portfolios
- to tap into existing portfolio projects
- to encourage people to build uniquely individual projects that serve their needs
- to create a well articulated, integral system
- to avoid getting bogged down in management issues

FOCUS: Portfolio assessment as a system for providing evidence for "valued learnings" which are not fully assessed through on-demand or curriculum embedded assessments

The definitions of "valued learnings" will evolve from the work of the Portfolio Task Force, but include abilities such as these:

- students evaluate their work, revise it and improve it over time
- students demonstrate problem solving

skills and explain the circumstances under which one method is better than another
- students demonstrate a willingness to take risks and challenge themselves

The task force intends to focus beyond management issues of what portfolios look like, onto questions about what we want students to learn and demonstrate and how teachers facilitate this accomplishment.

RESEARCH QUESTIONS:
Here are some questions that are being considered:

- valued learnings: What do we value?
- evidence: What is evidence of student accomplishment?
- equity: How do we make sure that the system and what it asks for is not biased?
- system development: How to develop an equitable, accessible system that provides important information?
- staff development: How do we involve large numbers of teachers in the development, piloting and scoring as assessment?
- scoring: Who, what where and when? (The hope is that portfolio evaluation doesn't get collapsed into a meaningless number...What does "individual score" mean?)
- validation: Comparability, reliability, validity of scores...
- state system: How do the three forms of assessment work together?

Karen posed the following questions and ideas which seem to be parallel between ETS/CLAS and NSP:
How can we best accommodate the great diversity of ideas about what a portfolio is or should be? We are taking a "first principles" approach. What are the most valuable kinds of information that we can get from portfolios? If we can focus on that, then people can put forward their models in terms of how well they get that information. This is a way for people to come to grips with refining their models. The most important thing is getting people to have conversations about that information while student work is in front of them. This is critical to any system of standards development.

Also important to understand is the idea that portfolios don't have to carry the whole weight of assessment.

When discussing portfolio issues, how do we accommodate what people already know how to do, and also move things along toward more refined models.

Whatever is cross-curricular (in the valued learnings, or in other aspects) ought to be drawn out, made the most of.

Pittsburgh's Portfolios

Joanne Eresh

Joanne outlined the writing portfolios in Pittsburgh and the value of cross-curricular work to one's own discipline...

History
Howard Gardner invited Pittsburgh to take part in Arts P.R.O.P.E.L.

We began by getting together with teachers (middle school and secondary), ETS consultants, Harvard folk. We asked, "What does it mean to be able to, for example, write poetry? What makes it good, bad?

We came to the idea of a writing portfolio...its predecessor in Pittsburgh was the writing folder which was a wasted resource. So we fell into the idea of compiling a portfolio with these goals in mind:

- students taking responsibility for learning
- bring about a marriage of instruction and assessment

How It Worked
Students collected work during the first semester. At midyear, the students pick 5 pieces:.

- first selection: "Find the most important piece you've written" (portfolio as a cause for conversation)
- second: "Choose your most satisfying piece of work."
- third: "Choose an unsatisfying piece." ("How would you make it satisfying?")
- fourth: A free pick...("Whatever completes a picture of you as a writer")
- fifth: A final reflection

The important thing about the portfolio is the creation of the portfolio. An activity that, at first, seemed almost mechanical, was an important learning experience. Down the line, choices from January no longer seemed the right ones. So students were free to choose again.

When we started looking at kids' writing, we didn't know what we were looking for (and it changed anyway—not to a lower or higher standard, but to one different from our prior expectations)

Questions posed to students:

- Why did you select this piece of writing?
- What are the special strengths of this work?
- What was especially important to you as you wrote this piece?
- What did you learn about writing from your work on this piece?
- What kind of writing would you like to do in the future?
- Identify a particular area to work on in future.
- Describe the most significant revision you made to this piece.

If you ask a student to describe the assignment, you find out what she perceives it to be, which may or not be what you perceive.

An outsider ethnographer, Roberta Kemp, took notes on teacher discussions of student work, and drew our attention to things we were saying consistently—things like: "Gee, this kid worked hard on this piece...," "This piece is funny..."

Issues
What we used to value: Spelling, grammar, punctuation was what was taught, or at least, corrected.

The kinds of evidence contained in student work:

- planning
- drafts
- revising
- publishing

Scoring:
Since the focus was the program, we did not give individual results, but rather, school results. Teachers related that they learned more about teaching writing from this process than from any other experience.

Reliability:
If teachers have opportunities to talk about growth, development, process and strategy, they can score reliably. When 5 teachers were brought into the process (new, but not unacquainted with the writing process) and gave a few hours of training with several exemplary portfolios and had people score within a grade level, they got over 95% agreement. The scoring was more reliable (and less inflated?) when scorers only read final drafts, because they were not unduly influenced by the degree of hard work.

There has to be room for the unexpected, but strong, piece of work. If we value certain kinds of work, we have to find a "scale" that allows (or encourages) the unanticipated response.

Students sometimes produced cryptic brainstorming/thought processes:
If you don't know what the "murderous dog" and the "chicken bone" mean, then ask the student.

Revision:
Why does revision sometimes make the work less good? (To avoid this, students must be revising for a purpose, not just to revise.)

Development as a writer:
Sometimes hard to see when first and last pieces are very different genres.

Reflections

Ruth Parker

Phil requested that Ruth Parker attend the meeting to reflect back on what was discussed during the meeting. Here are some of her thoughts:

History
She is currently working in Edmonds, Washington assisting teachers who took up portfolios last year, some of them reluctantly. First and second selections ran just ahead of parent meetings. It was only in March that it began to come together and emerge as a powerful process.

Question: *What do teachers want to protect from being messed up?*

1) **student selection:** opportunities to demonstrate their own uniqueness and to value that of others. Student selection allows us a view of how

students see themselves as mathematical thinkers.

2) teachers' reflection upon their instructional programs

What can portfolios do and get at, that other methods cannot?

1) students reflecting on their growth over time
2) group investigations
3) math disposition—going from answers to new questions

This is a process of discovery; we don't know what others will emerge

Ruth's portfolio sheet from last year—students choose:

one's most challenging piece of work
a task that showed growth

She hopes that we are able to avoid bringing in "portfolio" as a well-defined product that does not allow for building meaning.

She pleads considering broadening the "substantive exchange" beyond the student and teacher to include others.

Other questions/comments she has include:
Can we establish criteria that truly value students being challenged, struggling with ideas, taking a risk?

Can we engage in a process that lets us stay for a while with our questions before we define...?

Can portfolios continue to be defined as a process, not just a product?

Resnick says things that aren't assessed disappear from the curriculum. Portfolios may be a way to encourage some of the qualities we value:

persistence
curiosity

Are we inventing new labels that sound nicer to us, but still translate back to the child as "I'm good", "I'm not good"...?

Can the portfolio be assessable and also let students show the aspects of math that exemplify their strengths? (Mathematicians specialize; many students become convinced early on that math "is not their game").

There is tension between the concept of "lifelong learners" on one hand, and "scores" on the other...

Even though standards are essential; we have to be careful in selecting the standards we will embrace.
Mathematical language: be sure that we value it according to how students use it, rather than valuing it just for itself. Since mathematical language can mask misconceptions, it's important to look at how it is used.

We have to ask students what they are thinking in order to find out.

Need the best prototype lessons we can find, that teachers will want to use, and that will support the teachers' role as instructional decision makers in their classrooms.

Rather than working from scripted responses, teachers who "spread the word" need to have experiences in mathematical content, processes, etc. Teacher support (thinking of them and treating them as professionals) is essential to making this work. Teachers must have the same real and rich experiences that we want them to provide to students. We need new models other than "trainer of trainers."

Are we looking for both student assessment and program assessment? Can they be used for the same things? Must there be scores at the student level? Can assessment focus more on what students can do and how to improve their abilities, rather than detailing what they can't do?

<u>Follow-up discussion:</u>
f "trainer of trainer" models is not feasible, then what is?...there must be a feasible middle road between Ruth's ideal and the reality.
Ruth's worry that we are wearing out the "trainers." In reference to Marge's testimony about growth in teachers' understanding of the NCTM Standards during the past two years.

Is there a grassroots effort in Kentucky or Vermont toward having teachers writing tasks?
Beth Hulbert: "This has not been productive or effective in Vermont. We haven't found many who can tell us about a good task (how to write them)."
Marge Petit: "My best tasks come out of working with my students. In a staff development setting, it hasn't worked well."

Karen Sheingold : "But if you had a situation in which the task was not separated from instruction, maybe task creation would be more effective."
Sue Rigney: "this summer's training will focus on teachers seeing themselves as problem solvers and transmitting that to students. This is a need recently perceived by teachers (they've grown into an awareness of needing it)."
Ruth Parker: "It's hard to ask teachers with a traditional, arithmetic-based background to write effective tasks."

Concluding Comments

Phil Daro

This meeting was the worthwhile learning experience we all expected! We have gotten real and valuable input from the participants. This will translate into concrete building blocks for a successful portfolio assessment system. We, at the New Standards Project, are counting on your continued assistance in making this a reality.

For Further Information

Contact:

Sue Rigney
Vermont Department of Education
Planning and Research
120 State Street
Montpelier, VT 05602-2703

Marge Petit
RR1 - Box 1770
Plainfield, VT 05667

Beth Hulbert
Ward 5 School
Beckley Hill
Barre, VT 05641

Cheryl Tibbals
State Department of Education
Division of Performance Testing
500 Mero Street, Room 1926
Frankfort, KY 40601

Nanette Seago
6179 Oswego
Riverside, CA 92506

Barbara Storms
Educational Testing Service
Research and Development Division
1707 64th Street
Emeryville, CA 94608

Karen Sheingold
Educational Testing Service
Rosedale Road
Mail Stop LR
Princeton, NJ 08541

Joanne Eresh
West Liberty Center
Writing and Speaking Division
Dunster & Lamoine
Pittsburgh, PA 15226

Ruth Parker
4130 Stuart Circle
Ferndale, WA 98248

Phil Daro
New Standards Project
University of California
300 Lakeside Drive, 18th Floor
Oakland, CA 94612-3550

Issues in Scoring Cumulative Accomplishments
Implications for Portfolio Design

A PAPER PREPARED FOR THE NEW STANDARDS PROJECT

by Philip Daro

June, 1993

The purpose of this paper is to pose issues and suggest possible approaches to portfolio scoring based on standards. Many of these issues have existed for a long time in the arena of report card grades, but the habit of the system has been to leave grading standards and practices to individual teachers. The New Standards Project can be thought of as establishing a more broadly responsible and supportive context for the assessment of student accomplishments. Teachers, students and the public will be linked through national standards and an accountable process for assessing the cumulative accomplishments of students against the standards. How can this be done across the variety of partner states, schools, teachers and students?

The focus of this paper is the issues related to scoring cumulative accomplishments. These issues are organized in the following topics:

What to score: individual entries, aspects of performance, the collection: holistic vs. dimensional vs. analytic vs. hybrid

Criteria

Exemplification

Process

Selection of ingredients: who selects, what is required, standard tasks, types of entries, reflections, self assessment

Rights and responsibilities: student and teacher roles, NSP, Partners, public

On-demand performances

Difficulty, Weakness of Stimuli

Performance Standards Defined on a Body of Evidence in Submission

Many of the questions discussed below resolve differently for different purposes of assessment. Too much can be made of differences in purpose or uses of assessments; for all practical purposes, those responsible for designing and installing assessments cannot control the uses to which they will be put. Almost any scrutable assessment at the student level will have heavy consequences for at least some individual students in at least some schools, even if it 'only' influences the student's and parents' own estimation of the student's worth and potential as a student. Therefore, we have to design for this

use, even when it isn't part of our system. Likewise at teacher and school levels. In other words, in addressing the issues below, we have to anticipate consequences beyond our intentions, and, within reason, take design responsibility for the validity of such consequences.

What to Score

What are we trying to assess? Should we score tasks, selections of tasks, dimensions, aspects of performance, traits, or some combination? How does what we score relate to performance standards based on curriculum standards? How does what we score relate to what we prompt? What will our scores communicate to students and teachers? What if teachers emulated our scoring in their report card grading systems, would our scoring schemes serve them well? These and other questions will be resolved in the course of implementing NSP assessments over the next few years. Of immediate concern are the decisions related to cumulative accomplishments.

There is much discussion that contrasts holistic scoring with dimensional scoring. For scoring cumulative accomplishments, I believe this discussion can be restructured so a synthesis of the insights precious to each point of view can be achieved, and a more powerful and useful system devised. Each type of scoring solves a different problem by making different trade-offs.

Holistic

As used to score single performances, holistic scoring respects the wholeness and interconnectedness of the performance by refusing to impose any a priori analytic structure on the structure created by the student in performance. The holistic scorer is asked to judge how well the performance accomplished the purpose set in the prompt[1]. The scorer uses anchor exemplars and a criterion (rubric) that, together, express the prompted purpose as a scoring standard. The anchors show the variety and range of performances. This holistic judgment allows for great variety in how a student approaches the accomplishment of the purpose. In particular, it allows for the student to respond in whatever dimensionality he or she chooses.

For example, in assessing how well students can formulate simple mathematical models of realistic situations, students must create their own structures and variables (this is central to what we want to assess). Some may use geometric thinking, others will create functional relationships, while still others will overpower the situation with the cunning use of arithmetic, expressing generalizations in practical advice. While these approaches differ in their mathematics, they all accomplish the set purpose, and can be scored with a common holistic rubric.

There has been considerable success in scoring writing and open ended mathematics tasks using task specific rubrics and anchors. Most experienced scorers and teachers I have talked to doubt that a general rubric would work very well. Task specific rubrics can

---

1 Of course, the purpose that animates and steers the student in performance is constructed as an interpretation of the prompt. There will typically be many reasonable variants of performer's purpose, and therefore the purpose is only set in the prompt in a half baked way . Such variance in performer purpose raises a number of problems : By reasonable interpretation, a student can make a task more or less difficult. A second problem is to account for the variation in anchor papers and rubrics. The design and crafting of prompts can help mitigate these problems, but cannot eliminate them. The attempt to mitigate these problems often creates other problems, such as denaturing the task, pre--empting the opportunity to perform by too much leading, etc. Grappling with the effective communication of purpose is a key to the craft of performance task design. It requires many iterations of student trials.

deal with the purpose of the task in very direct, easily interpreted ways. Knowledge from pilot testing can be used directly in the construction of the rubric.

Many do think, however, that genre specific rubrics might work for different prompts within well understood genre. This has more immediate use in writing than in other subjects, although promising exploration of these ideas are underway in mathematics. If genre specific rubrics can be made to work, then teachers or curriculum leaders can select their own prompts within genre, even in a national system like NSP.

Holistic scoring of portfolios can be based on the holistic scores of each entry. One advantage of this approach is that we know the most about how to do this kind of scoring reliably. The scores are also directly applied to an object (the piece of work) that is real to the student, teacher, and public. On the other hand, the set of scores for the set of objects lacks much meaning of its own beyond *how well the student has done what he or she has been asked to do.* . Notice that this is approximately the same meaning as a report card grade. It may be possible to give considerable meaning to this kind of score set by also scoring what the student has been asked to do against curriculum standards; an opportunity to perform score!

Another approach is holistically scoring across entries. Kentucky is using this latter approach. A similar approach has been piloted on a small scale in California. When scoring across tasks, the possibility arises for **setting a purpose for the selection** of tasks; the purpose of the selection is greater than the sum of purposes of the individual tasks. This purpose can be communicated directly to the student and teacher making the selection. In this case, it functions as a higher order prompt for the portfolio as a whole. When this is done, students and teachers are often asked to comment reflectively on the selection vis a vis its purpose. The criteria used by the assessee in making the selection should parallel the criteria for scoring and the curriculum standards. Kentucky has made a deliberately transparent attempt to do this.

It may be useful to distinguish true holistic scoring that emphasizes the integrity of the performance and seeks to judge it on its own terms from *fused multi-dimensional* scoring. The Kentucky and California approaches to cross selection scoring use an *a priori* multi-dimensional schema based on curriculum standards. Scorers are asked to make one overall judgment of how the portfolio selections as whole meet the multi-dimensional standard as a whole. In some ways this multi dimensional fusion lacks the directness of true holistic scoring. On the other hand, it makes the standards for balance in the curriculum more explicitly assessed.

Holistic scoring relies heavily on the range and balance of tasks to link the scores to the curriculum standards. The scores connect to the curriculum through the tasks' connection to the curriculum. This raises the question of who is responsible for the range and balance of tasks: student, teacher, school, partner, NSP, professional community, instructional materials developers, others? The answer will surely involve all these people in some way. This issue will be taken up below. It is noted here, however, that holistic scoring depends more than other scoring methods on a satisfactory answer to this question, since the selection of tasks are such a central part of the definition of the performance standard. For this reason, holistic scoring may require more standardization of tasks than other methods.[2]

---

2 Every argument I have heard or imagined against this conclusion that holistic scoring leads to more standardization has been of the following three kinds:

The selection of entries as a representation of the curriculum -- the range, type and depth of assignments students are being taught to accomplish -- reflects the students' opportunity to learn in a direct way. This direct evidence does not tell the whole story, but it can contribute a necessary piece of the puzzle in the actual work of students. It is possible to obtain an opportunity to learn profile from portfolio scoring (see below).

Some public explanation of how the portfolios relate to performance standards rooted in the curriculum standards will be needed. It will be easy to provide vivid exemplifications, but difficult to explain the breath and detail of what is systematically assessed.

Dimensional

Dimensions[3] derived rationally from the curriculum standards can be formulated so scorers can evaluate how well a performance exhibits power in a particular dimension. For example, Vermont is interested in how well students generalize and make connections to other mathematics beyond the solution to a problem (see NCTM Standards, especially 4). They score a dimension referred to as the "So What" dimension. To score high, a performance must go beyond the solution to a generalization. This is true for all tasks. Other dimensions used in Vermont relate to use of mathematical language, approaches to problems, and other curricular goals originating in the NCTM Standards.

As with holistic scoring, dimensional scoring can be applied to individual entries (Vermont) or across entries to the selection as a whole.

A weakness of dimensional scoring is the necessity for assuming we know the dimensionality of the performance. The move toward more realistic curricula and pedagogues calls for more realistic assessments. The dimensionality of realistic performances varies much more interactively among individual by context by task by interpretive assumptions of the assessee by personal factors (just like reality does) than artificial performances. The same scoring dimensions applied to performances of widely varied dimensionality will produce measurement errors that correlate highly with the dimensionality of the performance. This may be unfair in gross ways to

---

1. External, standards based assessment shouldn't impose on the reality of teacher choices and therefore all arguments that go back to standards are invalid.

2. The link to national standards is rhetorical anyway, so it isn't as important as the integrity of the teacher student relationship to their own circumstances. Only at they can decide what is right for them. Therefore it is OK for the link to standards to be very general.

3. Types of tasks can be partially standardized without standardizing the tasks themselves.

Arguments 1 and 2 are not available to NSP, whatever their merits. Argument 3 has possibilities, especially in a system that includes an on demand exam in complement to portfolios. Kentucky is trying this.

3 Lee Cronbach has suggested that this use of 'dimension' can be misleading. Clearly, the dimensions used in scoring are not the factor analytic dimensions that come to mind in an assessment context. Intercorrelations are often very high. "Aspects of performance" is more accurate, if inconvenient. I assume we will come up with a better term, at which time I will replace dimension throughout the text.

individuals or classes of individuals, if there are consequences to individuals or classes of individuals.

One strategy for addressing this weakness also adds an important strength. **Teach the dimensions used for scoring to the students** prior to the performance. To the extent they learn to self-assess in the desired dimensions, their performance will be in the dimensions used in scoring. To the extent that the dimensions are central curricular insights, learning to use them will be of central curricular value. Thus, dimensional scoring with a strong self-assessment component amplifies the curricular influence of the assessment. Teachers in Vermont and Oregon are enthusiastic about the effects of teaching dimensional self assessment to their students as young as fourth grade.

How well the central values of the curriculum can be expressed as scoring dimensions for use in self assessment is an intellectual and empirical question on which the soundness of this approach depends. Criteria like 'so what'. really communicate a piece of work, a generalization in this case. Other criteria are more like the advice given to the sprinter, "run faster". While running faster certainly sums up the ultimate performance dimension, it is uninformative to the sprinter and lacks utility for self assessment. "Work on your mechanics" is barely any better. For the self-assessment argument to hold as a rationale for dimensional scoring, the dimensions have to inform students of how they might improve performance.

The 'so what' type of criterion does this because it denotes a ingredient that can be added by the student, criteria that look at the quality are more elusive for the student. This question may answer differently in different areas of curriculum. Expressing curricular values as a small set of dimensions risks encouraging simple minded versions of curriculum that invest too much in a few tricks that have large payoffs in the simple dimensionality of the scoring scheme. Holistic scoring represents curriculum as " tasks like these: exemplars of tasks". The meaning of "like these" is left to the various interpreters, including the student. Dimensional scoring provides a framework for interpreting the tasks in a general way, giving some definition to "like these".

Scoring Collections: Selection * Dimension

The table below frames possibilities for scoring a collection of student selections. The columns represent individual selections, while the rows represent aspects of performance (dimensions).

	Selection #1	Selection #2	Selection	Selection	Selection ...#n	
Aspect A						Aspect A Score
Aspect B		Critique and Self Analysis				Aspect B Score
Aspect C			Critique and Self Analysis			Aspect C Score
Aspect ...				Critique and Self Analysis		Aspect Score
Aspect ...						Aspect Score
	Holistic Score	Holistic Score	Holistic Score	Holistic Score	Holistic Score	

Aspect scores compare evidence from a selection of performances to criteria based on national standards: *How well do the selections evidence problem-solving power? power to use multiple representations effectively?*

It is possible to obtain row scores directly by having judges evaluate the collection as a whole. One suggestion for this procedure is to train judges to search for evidence like a detective, finding it wherever it happens to be located. A student could get a high Aspect B score for an impressive body of evidence distributed throughout the selections or concentrated in just one selection even though other selections lacked evidence.

Alternatively, row scores can be obtained by scoring each selection on Aspect B, and then aggregating across the row in some fashion. The disadvantage of this kind of procedure is that it enforces a stereotyped dimensionality on all selections. Unless the selections were intentionally all of the same genre of work (expository prose, for example), this would likely be reflected in an unwanted monotony and narrowness in the curriculum, no

matter how clever the Aspects were formulated. The Aspects will have a broader, more satisfying meaning in rhetoric than they will in practice. The practical meanings are inevitably more mechanical and banal.

One way around this difficulty is to use a balanced list of aspects from which different subsets would be applied to different tasks. The procedures for deciding which aspects to apply to which selections raises interesting possibilities. The designers of a particular task could be asked, as part of the design, to articulate the aspects which apply. Teachers, students and scorers would all be using the same scorecard as they played their respective roles. Yet even with a balanced list of aspects, this promotion of the scorecard to the foreground will tend toward simple-mindedness of curriculum. How long would the list of aspects have to be to avoid this "simpling" effect on the curriculum? As long as the NCTM standards, perhaps?

A variant of the selected aspects approach is to let the chips fall where they may; ask the judges to select aspects to apply to a selection on the basis of the evidence before them (the actual student performance). To be fair, this approach would probably have to use a 'positive evidence' method. An aspect would only be triggered for application by the positive presence of evidence. The lack of evidence would not contribute to a selection score. In this variant, some determination would probably have to be made about how to interpret the lack of evidence for Aspect B, for example, across an entire row. The burden can be placed on the student and teacher to submit a balance of selections that supplies ample evidence of all aspects. With the burden of evidence so assigned, the lack of evidence across a row can be interpreted fairly as evidence of lack in performance.

Column scores can be obtained directly through holistic scoring of each selection. **Holistic scores** compare performance on a selection to the purposes of the selection: *How well did the student accomplish what he or she was asked to accomplish?* This corresponds to teacher grades, and report cards.

Whatever aspects relate to a selection are expressed in the prompt and criteria for holistic scoring. The holistic score can be supplemented by a small number of additional scores for aspects (often called traits, in this context). These scored aspects can be drawn from the balanced list based on the genre of the selection, or be specific to the selection.

Column scores can be obtained by aggregation of scores for each aspect applied to the selection. This amounts to an analytic procedure. It requires a great deal of presumption regarding how the aspects relate to a particular student's performance.

Cells have no generalizable interpretation (no generalizable score)[4] across columns, rows or students. The variation in dimensionality from performance to performance(inclusive of all possible dimensionalities) is too much a matter of performer choice for any performance realistic enough to be qualify as realistic. Anything we do to eliminate choices, eliminates the possibilities for assessing the students' power to make effective choices. We are left with the anemic domain of the

---

[4] The relationship between an individual cell and its row, or aspect, is very unstable because for any particular selection, the importance of the aspect vary wildly from student to student depending on the strategy they used accomplish the task. The dimensionality of the selection is determined to an important extent by the student. The relationship between a cell and its column will be unstable because students will rely on different aspects of mathematical power for the same task. This is without even worrying about the differences among selections across student. For an individual student, however, the cells have meaning during instruction as revision tools.

assessors' predetermined choices; we can only assess how well the assessee can guess our choices.

This partly explains why cell scores may not generalize well even within student across task. The same student may very legitimately choose different dimensionalities of response for tasks that appear to the assessors to be very similar. Indeed, they may be motivated to do so by the assessment situation itself.

Nonetheless, cells are useful analytic categories for student self assessment and teacher critiques during instruction. For these purposes, the systematic consideration of each aspect for each piece of work raises useful issues that could lead to producing higher quality work, and learning how to produce higher quality works. To raise issues for consideration is one thing, to decide fate is another; cellular analysis is a questionable procedure for determining grades and other high stakes assessments.

Genre

Using the table above as the conceptual base, higher order concepts can be used to organize assessment designs. Two of the most promising are genre and syndromes.

Selections (the columns) can be assigned to genre of similar selections. In writing this has been a common and useful practice: persuasive essay, autobiographical sketch, expository essay, poem, story, descriptive report, etc. In California, a taxonomy of writing genre is explicit as the basis for the writing sample assessment. The taxonomy is much broader than the genre traditionally taught. A direct consequence has been the opening up of the writing curriculum to include more genre.

In mathematics, the genre of work in traditional programs is radically narrow. NCTM and others have called for a serious broadening in the kinds of assignments given to students. It may be possible to express this curricular goal as a balance of genre, although this is not yet clear. Other subjects will probably be easier than mathematics, but not as natural as writing.

Genre, or something like them, can solve a number of problems. Each genre could have genre specific standards that derive from the purposes of the genre (persuasive essay has persuasion as its central purpose). These purposes transcend particular prompts within the genre. There are four related advantages to this:

1. Students can learn how to perform in a genre, transferring across performances within genre; they can get better at working in a particular genre. Genre level assessment criteria can be powerful self-assessment tools. Students can learn from performances on different prompts within a genre. For example, students can study performances of other students on other tasks that exemplify the genre, as preparation for work on a new prompt in this genre. By developing a sense of genre, students also acquire the values of good genre performance.

2. Scorers can learn to score a genre. Rubrics can be constructed at the genre level. This does not preclude task specific rubrics within genre, indeed, it provides a common framework for constructing task specific rubrics. A genre score can be derived from holistic task scores within a genre.

3. Genre level rubrics can be used by external assessment systems like NSP, thereby allowing local variation in specific prompts within genre. This is especially important if we want to assess a healthy amount of curriculum; the alternative to local flexibility would intrude on too much instructional time.

4. In assessing cumulative accomplishments, some requirements can be made at the genre level: for example, 'include a selection that belongs to the genre *applied research report* (defined elsewhere) , the rubric for assessing *applied research reports* is provided along with exemplars at several levels of performance. Suggestions and sources of suggestions for *applied research report* assignments are also given.

Most genre can be defined in ways that make sense across grade spans. Standards for performance in a genre can be independent of grade level or can be specific to grade level. If we want to build in standards for the use of mathematical representations in *applied research reports* , we would probably want to tailor them to grade level expectations.

Below the table has been augmented to show where genre fit in.

	____Gen	re I___	_____Ge	nre II__		
	Selection #1	Selection #2	Selection	Selection.	Selection ...#n	
Aspect A						Aspect A Score
Aspect B		Critique and Self Analysis				Aspect B Score
Aspect C			Critique and Self Analysis			Aspect C Score
	Holistic Score	Holistic Score	Holistic Score	Holistic Score	Holistic Score	
	Gen Sco	re I re	Ge Sc	nre II ore		

Each genre can have associated with it a characteristic subset of dimensions, or aspects, that are particularly valued for that genre. Students can be taught these as in the Vermont example. The variation in dimensionality from response to response for a student within genre will probably be much less if the student understands the genre.

By varying dimensions from genre to genre, the simpling effect of dimensional scoring on the curriculum can be greatly mitigated. "So What" from Vermont works well in the *problem* genre that Vermont has used, but different dimensions would work better for *pure investigations*. It is also possible to give more specific and meaningful definition and exemplification of a dimension for the limited range of work within a genre than across all sorts of work. There is no reason why the same dimension cannot be defined and exemplified differently in two or more different genre. For example, a dimension might be 'mathematical reasoning' from the NCTM standards. In an *applied research report* the reasoning would be embedded in the application, and the constituents of argumentation would mix realistic issues with mathematical structures. In a *pure investigation*, the reasoning would adhere more closely to the standards for an abstract proof. Yet, both are expressions of a common dimension: mathematical reasoning as defined in the curricular goals.

The question of how the use of a higher order category like 'genre' affects dimensional scoring needs further investigation. It seems likely, however, that a meaningful organization of genre that captures important breath (across genre) issues of the curriculum while allowing for more attention to depth (within genre) can only help stabilize the dimensionalities of performance. This will greatly improve prospects for generalizability.

Syndromes and Profiles

Turning now to the rows, it possible to construct higher order patterns connecting dimensions, or aspects. Many have suggested the use of 'profiles' across aspects as interpretative devices. The intent is to preserve information and add to the utility of the assessment. Such suggestions often evoke images of diagnostic profiles. But a profile merely presents the dimensional scores as a set. The interpretative process, whether employing some form of cluster analysis or not, imposes a post hoc theoretical pattern on observed data. This can be useful, but we should not settle for it. If the theoretical patterns have merit, let us put them to the test.

A priori patterns across assessment profiles might better be called syndromes than profiles. This term correctly suggests our obligation to connect the patterns to causality and also to a consequential response.

The dimensions relate to the particular construction by a student of a response, a construction involving choices as well as knowledge and mathematical power. If students are well informed about the dimensionality by which the response will be judged, then it is fair to let the score depend on the choices. The choices become part of what is assessed and part of the curriculum. If not, it is not fair.

The dimensions also relate to the curriculum. We are assessing performance in a curriculum. Creating dimensions that express the values of a curriculum is necessary, otherwise students and teachers are led away from the curriculum toward what counts. The eternal question, 'Is it gonna be on the test?' exemplifies how this connection operates in report card grading. What craft and technical know how do teachers employ to express their own curricular values in their report card grading schemes? How do these compare with their colleagues and national or local standards?

Since the dimensions can be standard and somewhat independent of particular tasks, especially if genre dependencies are established, the scoring can be to a standard that is not undermined by weak assignments. It would not be fair to a student to give him or her a low score on a dimension on which he or she had no opportunity to perform (because of

weak assignments). But it is very much to the point to give the program (classroom or school) a low score on dimensions where their students performed poorly for whatever reason.

Syndromes can be used to characterize how different actual curricula in practice compare to standards based curricula. In the NCTM Standards, an explicit comparison is made identifying the shifts in emphasis from the traditional curriculum to the standards based curriculum. From this, a traditional 'syndrome' can be defined and looked for in the score patterns of schools to identify curricular opportunities to learn.

Opportunity to Learn

There are, of course, many factors contributing to a students opportunity to learn. Some are more direct than others. Among the most direct and specific are those that can be appraised in portfolios of student work at classroom and school levels. A classroom sample, or school sample of portfolios can be appraised from this standpoint.

First, an appraiser can readily determine the breath of assignments included in the portfolios. Are the genres that the curriculum calls for well represented in every students portfolio? Are some groups of students working in a good balance of genre, while other groups are working in a constricted set (low order skills, for example)? What is the pattern of performance across dimensions? Is one or another syndrome of ill balanced curriculum in evidence for all or some groups of students?

Such appraisals can be made for programs quite readily and reliably. Indeed these judgments have proven to be easier than the judgments about students. The students performance is contingent upon an opportunity to perform. When, as in a portfolio, performance is embedded in the curriculum over a substantial time sample, then opportunity to perform converges on opportunity to learn. Students with narrow opportunities to perform can still score well on tasks they did perform, but their scores on dimensions or aspects will be low. Scores on some genre are also likely to be low, since some genre are probably going to be missing or superficially represented.

The interpretation created in reporting such results must, in fairness to the student, make clear that the student did well what they were asked to do. What they were asked to do, however, lacked balance in specific ways and therefore scores for some dimensions and genre are low. The consequences of the assessments can be more validly aimed in this way. If the student does not do well what they have been asked, it is one thing; but if they haven't been asked it is another. Whoever manages the curricular priorities is responsible for the balance of genre and dimensions.

The quality of instruction that prepares students for their assignments is another matter. The distribution of selections across genre, and the fullness of their dimensionality can be appraised somewhat independently of the quality of performance. But the distribution of performance scores must be considered in order to support any inferences about the quality of instruction. These are dangerous inferences. Many, including the student, are responsible for the quality of performances. Nonetheless, the distribution of performances across students over years should be part of a larger body of evidence supporting judgments about the quality of instruction as part of the opportunity to learn.

Difficulty

A closely related issue is the difficulty of the challenge inherent in the selections. Variation in the challenge difficulty from student to student, class to class and school to school undermines the fairness of comparing scores. This problem can be greatly improved by communicating the appropriate challenge to the students and others making the assignments. One way to do this is through standard tasks, or standard exemplars of tasks along with commentary. Even better, along with standard exemplars, is to set up a system whereby students can query a reliable authority on assessing challenge difficulty regarding a task of interest. Such feedback can have a strong moderating effect on the volatility of standards in the system.

Clearly, the best resource to develop into such an authoritative feedback system is the community of teachers and students. If they can give reliable feedback to each other across classes and schools and states, the comparability problems arising from student and teacher selection of entries are much less. Even when they make bad selections, it is, to a great extent, their right and responsibility.

It would be convenient to employ normative methods for determining difficulty; but I do not think the inferences that can be drawn from normatively established scales of difficulty have consequential validity in the situations for which external assessors are primarily responsible: accountability, formative and summative evaluations. Normative methods inherently confound the loci of responsibility and consequence by ignoring the differences in cause of task difficulty.[5]

Is the source of difficulty the quantity or quality of opportunity to learn? Is it inherent in what is being assessed (understanding of a difficult concept)? Is it inherent in the design of the task (difficult problem, but requiring ordinary mathematics)? Is it cultural interference between the background of the student and the background of the task? Is it in the circumstances surrounding the performance (easy task, but not enough time)? By confounding these and other sources of difficulty, normative methods invalidate distinctions needed to identify responsibility and take action to improve future performance. We certainly need to distinguish opportunity to learn from conceptual difficulty, for example.

Such distinctions are particularly critical for students, teachers, and local leaders trying to improve things. Assessments that cover up these distinctions can and often do bolster existing beliefs and prejudices about who and what is responsible. We need

---

5 Effects due to student abilities are confounded with student effort, and these with opportunity to learn, and these with every other input variable for which someone should be responsible. Normative methods, in general have ill effects on systemic and consequential validity. Both validities are grounded in responsibilities: who is responsible for causing the condition being assessed, and what are they going to do about the assessed condition, in consequence?
Normative methods, by confounding causes (of difficulty, for example), insulate those with the power to cause from attribution, and thus from responsibility. Under these circumstances, the negative consequences tend to settle to the lowest levels: the student and his or her "background" (can we disown our backgrounds? should we? what is that part of me that is not related to my background?). Positive consequences tend to be shared by all levels.
What we need are methods that distinguish opportunity and accomplishment from background and ability. Assessments heavily influenced by components for which no one can be held responsible have little legitimate use as accountability tools at any level. Intentionally or not, such instruments have the effect of covering up responsibility and breeding a quasi-factual basis for fatalistic attitudes towards the effects of education. This is a sad irony, given that a deep purpose of education is precisely to overcome the hopeless fatalism of stagnant social orders of inherited opportunities.

assessments that highlight the causes of performance that someone can influence: the student, teacher, curriculum , school structure, or community.

Students and teachers need a curricular basis for evaluating difficulty. We need challenge appraisal tools that would allow for situations where most students succeed at something appraised very difficult (per the curriculum), and also where most students fail at something very easy, but rarely taught.

Ice skating is not a good analogy, nor is diving: in these cases the curriculum is radically narrow, defined almost entirely in terms of narrow performance skills, and defined only for distinctions among the hyper elite performers. The skating scales as scorings are simply misleading and invalid for beginning skaters, novice skaters and even proficient skaters (neighborhood hockey, Friday night at the local ice rink, I'll race you across the pond, etc.). No secondary culture of informal evaluation has ever trickled down from the hyper-elite scoring to influence evaluations of 99 % of the people who skate. The most notorious trickle from the Olympic schemes has been their use to rate women in sexist conversation. I am afraid any system that emulates scoring schemes that work fine within hyper-elite performances will degenerate into tools for stereotyping.

If standards are to have any real consequence, it will have to be through the engagement of teachers in a professional community holding each other to a mutually accountable standard. They can only hold each other to standards they understand in terms of their own students work. Thus, deliberating their students' work with colleagues in open but moderated scoring discussions will be needed to make standards a reality for teachers and thereby, students.

Grades and Report Cards

The United States has long had high stakes assessment at the individual student level: grades and report cards. A students future opportunities are profoundly affected by report card performance, even in the primary grades. The validity of these assessments for their consequences to the students has never been properly evaluated. Worse, serious efforts to improve the validity and reliability of these are rare and local. The professional community has taken little responsibility for the technical quality of the practice of its members in this fateful area. Individual teachers have had nowhere to turn for guidance and standards; they have been left on their own.

From a systems standpoint, grades are of almost no use above the individual level for assessing performance at classroom, school, or higher levels. Even if the problems of comparability could be solved, a system that delegates everything to individuals who have a direct interest in the outcomes would have no credibility. Past practice delegates virtually everything to individual teachers: setting standards, scoring procedures, designing assessment instruments, scoring performances, due process, complaints, proctoring, auditing, recording, and so on. The interaction between new assessment systems and grading practices has the potential to profoundly alter the role of the professional community in its relationships with the public.

Developments in holistic scoring, especially widespread teacher participation in moderated scoring sessions can have a revolutionary impact on the quality of assessment for grades and report cards. Since grades and report cards are the operational expression of the value system of the teacher and school after grade 5, reforms will directly transform the value system, and with it the culture.

New Standards Takes Close Look At Portfolios

Design, Technical Rigor Highlight Discussion

As part of a national effort to develop performance-based assessment, a dozen portfolio experts met in Minneapolis in August to review the state-of-the-art knowledge about the effective use of portfolios in the classroom. Sponsored by the New Standards Project, this meeting is one of the first national meetings to consider portfolios as the centerpiece of district and statewide assessment. The meeting was organized for the New Standards Project by the National Council of Teachers of English and the University of Illinois.

Portfolios Widespread and Diverse

Since the mid-eighties, when education reformers turned toward portfolios to monitor and evaluate student progress, the portfolio movement has gained tremendous popularity in school districts throughout the country. As Stanford University researcher Robert Calfee observed, however, there are a variety of approaches to portfolio use.

Calfee, reporting on his national survey of teachers who use portfolios, reviewed how teachers initiate portfolio usage, why they incorporate portfolios in their classrooms, how the process of portfolio implementation works, and how teachers use the results.

Calfee and his colleagues found that teachers are "outrageously enthusiastic" about portfolios. "Especially," Calfee said, "if [portfolios] are not promoted from on high, but are something [teachers] pick up on their own, at workshops. They will put enormous energy and time in exploring the possibilities."

Calfee also reported finding a total lack of any technical rigor. "Issues of validity, reliability, and so on have simply fallen through the cracks," Calfee said.

New Standards Meeting Focuses on Portfolios

Continued from Page 1

Some portfolio projects have attempted to address technical rigor, however. Willa Spicer of the South Brunswick, New Jersey, school district reported that she had been unhappy with the district's portfolio projects even after several years of development because she had no way of testing the reliability of the information teachers were collecting from the portfolios. In some cases, she had no method of aggregation. "I didn't know whether one teacher was judging a child differently than another; I didn't know whether bias existed," Spicer said. That's when she and Ted Chittenden of Educational Testing Service developed a scale for districtwide use.

John Ford Photo

Jackie Cheong and Geof Hewitt

Chittenden said the South Brunswick scale is theory-referenced and developmental in nature. The scale describes what a child can do, and he said, "It has become increasingly useful for district decision making, for policy decisions, for monitoring and accountability."

Though Spicer and Chittenden reported high inter-rater reliability for their scale, Spicer said the most important part of portfolio assessment occurs after the numerical rating, when teachers get together to negotiate differences among their perceptions of the portfolios. "The important thing," Spicer said, "is you have a process here where, in a peculiar way, the task of putting a child on the scale has focused the teachers intently on the records of the classroom, and generated really good discussions."

> *You have a process here where . . . the task of putting a child on the scale has focused the teachers intently on the records of the classroom, and generated really good discussions.*
>
> —Willa Spicer

Portfolios Are "Low-Stakes" Tools

Geof Hewitt of the Vermont Writing Project cautioned about getting too involved in reliability. "We're striving so hard for reliability that we're forgetting what really happens in the classroom," Hewitt said. "It's that very personal, very idiosyncratic, very anecdotal exchange that the portfolio provides so well for, and when we try to reduce it to numbers, I think we get into trouble." Hewitt believes the reason for using the portfolio as an assessment tool is "to give the legislature data and to feed the newspapers." He said school administrators need to be reminded of the low-stakes nature of portfolio projects, so that teachers are not censured or rewarded on the basis of judgments made about portfolios.

One feature of the Vermont portfolio project that Hewitt considers among the most interesting is the inclusion of what the student believes to be his or her best piece, along with a reflective letter about the best piece. He thinks if students receive only a global assessment on their entire portfolio, it isn't very clear to them how the standards of performance have been applied. "But if a specific piece has been assessed, *and* the portfolio," Hewitt said, "the student can focus on that specific piece and perhaps come to understand how those standards relate to the larger [body] of work."

In the Pittsburgh public schools, students aren't asked to choose their *best* piece but rather the pieces they see as most important, satisfying, or unsatisfying, explained Joanne Eresh, school support service specialist in charge of assessment in Pittsburgh. The reason for incorporating portfolios in Pittsburgh was to provide students with a process for looking at their work and for making judgments about it, she said, in order to give students more responsibility for their own learning.

"We're also looking for a way to help students and, therefore, teachers to begin to see that learning wasn't something that happened on Tuesday, but was an accumulation of what went on before," said Eresh.

Involving Students in Their Own Assessment

This idea of students revisiting their work in the interest of learning was part of the reason for instituting portfolios at Humanities High School in New York, said teacher Gertrude Karabas. "Students will achieve if they write, if they rewrite, if they own their work," she said. The use of portfolios to improve student achievement also paid off in enhanced performance on state tests, which made sense to Karabas.

John Ford Photo

Gertrude Karabas

"It happened very naturally because kids had a great sense of themselves as learners and writers," she explained. The portfolio project at Karabas's high school was also intended to promote student involvement in their own assessment, a purpose shared by a portfolio project in New York and Connecticut described by Alan Purves of the National Research Center on Literature Teaching and Learning (NRCLTL).

BEST COPY AVAILABLE

Teachers involved in NRCLTL's portfolio project began their year by sharing their evaluation criteria with their students. Teachers then negotiated individual objectives and goals with each student. Students know what the minimum criteria are for an average grade, Purves said, and they know that to achieve a higher grade, they must show evidence of effort and improvement in their portfolio.

"We compare portfolio assessment to the courtroom dramas we see on television," Purves said. "Although in reality [teachers] assume a dual role as both judge and legal assistant, helping [students] to locate and to establish the proof for their case, it's up to the students to lay the evidence not only before the teacher but before themselves."

Knowing that they are in control of the "evidence" that goes before the "court," Purves said, is a powerful motivator for students.

Teachers Learn from Portfolios, Too

Purves also believes portfolio projects provide teachers with an opportunity for self-exploration, a belief shared by Jackie Cheong of the University of California at Davis. Cheong was involved in a pilot study of the California Learning Record (CLR), which among other things, considered whether the quality of teacher observations and judgments could be improved to meet rigorous demands for objectivity, consistency, and validity. Teachers found the anecdotal record keeping involved in the CLR led them to interact with students and parents in ways they had not done so previously. It also led them to examine their assumptions about how students learn, Cheong reported, and about the relationship of teaching practices to students' learning.

"[The anecdotal record keeping] also allowed them freedom to experiment with new strategies," Cheong stated, "and it led teachers to decide what the next steps for instruction were."

In discussing the NRCLTL portfolio project, Alan Purves said starting with an explication of criteria was the hardest part for teachers. Kate Jamentz, director of the California Assessment Collaborative (CAC), said teachers in one of the 29 projects in alternative assessments that CAC monitors found it helpful to begin *their* portfolio project by answering the question, "What is it you care enough about to assess?"

Jamentz referred to this as the "values exercise." Teachers begin to use content standards in English language arts by answering the question, "What knowledge, skills, and habits in writing do you care about in students as readers, writers, listeners, and speakers?"

Jamentz said, "There's a huge gap between adopting standards and being able to use them for instructional planning. Teachers are saying, 'I need to put this in my own words; I need to figure out what it means, when you say, criticize a piece of text or, challenge a text.'" Unless teachers understand the standards themselves, Jamentz said, they can't address them instructionally with students.

Discussions Will Continue

The portfolio experts at the August New Standards Project meeting reported that the majority of teachers with whom they worked said portfolios involved the hardest work they'd ever done, but the most rewarding. Sheila Valencia of the University of Washington said this is why there is so much variability in portfolio design—because teachers are working to adapt portfolios to their own classrooms. She does not believe this is necessarily a bad thing. "I think that when we think about portfolios for a project like New Standards, we need to think about how portfolios contribute to the larger system, and not whether or not portfolios should be providing the same kind of information that other indicators are providing," Valencia said. "I think [portfolios] have a unique contribution to make. I think they've made dramatic changes in what teachers do."

The Literacy Unit of the New Standards Project continued its discussions of what portfolios add to classroom assessments; what role portfolios play in the professionalization of teaching; and how equity in portfolio design and implementation can be addressed at an October meeting in Boston.

—A.F.

BEST COPY AVAILABLE

New Standards Portfolio Development Teams

Elementary School Mathematics

Vermont
Balanced Assessment (California)
Oregon
Portfolio Project (Washington)
Integrated Portfolio (Kentucky)
CAMS/Project Zero (Massachusetts)
Texas
Florida

Middle School Mathematics

CAMS (California)
California PACE
Math REN
San Diego PACE
Integrated Portfolio (Kentucky)
South Carolina
Rochester

High School Mathematics

Vermont
Texas
Pittsburgh
Balanced Assessment (California)
Complex Instruction (California)
IMP (California)
IMP (Colorado)

Elementary School English Language Arts

Peconic Teacher Center (New York)
Kentucky
New Brunswick, New Jersey
Oregon
Project Zero
Coalition Site School (Croaton)

Middle School English Language Arts

Harvard PACE
Central Park East (New York City)
San Diego, California
Vermont
Monroe Middle School (Rochester)
Bellvue, Washington
Applied Learning Academy (Fort Worth)

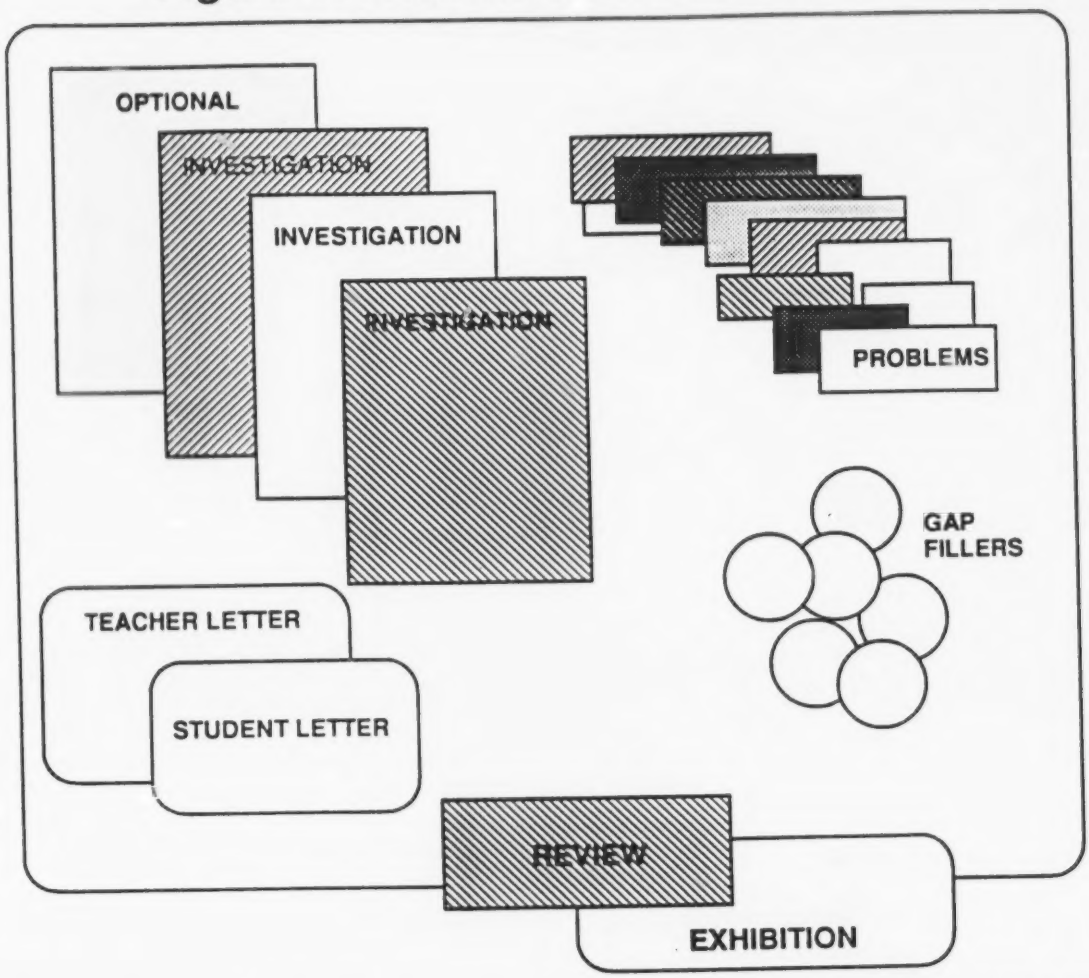
High School English Language Arts

Langley High School (Pittsburgh)
Horizon High School (Colorado)
Central Park East (New York City)
ROPE
Center for Writing and Learning

Cross Grade English Language Arts

California
Iowa

Figure 1: Mathematics Portfolio



INVESTIGATIONS: Showing in-depth, extended work; a project performed over a period of weeks using mathematics to research or design something practical, or investigating mathematical issues, or mathematically-powered interdisciplinary work. Combines group and individual work, and various methods of communicating results.

PROBLEMS: Showing problem solving, communication, reasoning, and a range of mathematical techniques and ideas at work.

GAP FILLERS: Showing the framework for balance has been covered; selections included included so the student can fill major gaps in the curriculum not addressed above.

STUDENT LETTER: The student's reflections on his or her work.

TEACHER LETTER: The teacher's comments on the student's work.

REVIEW: Verifying authorship and understanding of mathematical content, ability to explain and elaborate results, and communicate thinking to others. Applies to the portfolio as a whole, with special emphasis on investigations.

EXHIBITION: Making standards public to students, teachers, parents, and the community.

Figure 2: Mathematics Profile